Proceedings of CONF-MSS 2026 Symposium: Mechanical Control and Automation
DOI: 10.54254/2755-2721/2026.BJ31086

# *The Applications of Artificial Intelligence in Fault Data Collection and Fault Analysis*

**Yiyang Hu**

*Qixin College, Zhejiang Sci-Tech University, Hangzhou, China*
*huyiyangstu@163.com*

***Abstract.*** Against the backdrop of Industry 4.0, the intelligent transformation of industrial production accelerates. Equipment failures and process anomalies during production are likely to cause significant losses, while traditional fault management models have failed to meet modern demands. Artificial intelligence (AI) technology has achieved phased progress in fault data collection and analysis. This paper adopts a systematic literature review method to analyze relevant studies and cases, focusing on the application of AI in fault data collection and analysis. It aims to improve the research framework in this field, provide references for the application of AI technology in production, and thus enhance industrial safety and efficiency. This study identified key challenges, including data quality issues, difficulties in data sharing, and imbalanced dataset classification, as well as high costs and insufficient accuracy of data annotation. To address these problems, technical solutions such as data cleaning, federated learning, and resampling methods were proposed. In addition, it was recommended to adopt active learning and semi-supervised learning to reduce annotation costs and improve model performance. Looking to the future, the integration of generative AI and digital twins is expected to further overcome the problem of data scarcity, while self-evolving AI systems will drive the realization of more autonomous and accurate predictive maintenance. This research provides theoretical and practical references for the intelligent development of industrial fault management.

***Keywords:*** Artificial Intelligence, Fault Data Collection, Fault Analysis

## 1. Introduction

Under the wave of Industry 4.0, global industrial production is accelerating its transformation towards intelligence. The widespread application of automated production lines, intelligent equipment, and Internet of Things systems has not only significantly improved production efficiency and capacity scale but also led to a substantial increase in the complexity of industrial systems. Equipment failures and process anomalies have become core triggers for production interruptions, safety accidents, and huge economic losses.

Traditional fault management, which focuses on post-failure maintenance and regular preventive maintenance, can no longer meet the needs of modern industry. Post-failure maintenance requires troubleshooting and repair only after a fault occurs, which not only results in long downtime but may also cause fault propagation. Regular preventive maintenance relies on fixed-cycle planning,

DOI: 10.54254/2755-2721/2026.BJ31086

which tends to lead to over-maintenance or insufficient maintenance. More importantly, traditional fault detection and diagnosis are highly dependent on the experience of technical personnel. When dealing with hidden faults and multi-factor coupled faults of complex equipment, the identification accuracy is low, the response speed is slow, and the results are affected by subjective factors, making it difficult to meet the high requirements of industrial production for stability and safety.

Against this backdrop, artificial intelligence technology, with its powerful capabilities in data processing, feature extraction, and pattern recognition, provides a new path for the innovation of fault management and has achieved phased breakthroughs in the field of fault data collection and analysis. Through multi-dimensional, full-cycle operation data collected by sensors, industrial cameras, IoT terminals, and other devices, after preprocessing by big data technology, machine learning, deep learning, and other algorithm models can realize accurate fault detection, intelligent diagnosis, and early prediction, and output executable decision suggestions, effectively making up for the shortcomings of traditional methods.

This paper adopts a systematic review method to comprehensively sort out relevant literature and practical cases, and deeply analyze the application logic and implementation effects of core technologies, aiming to improve the research framework in this field, clarify the key points and optimization directions of technical application, provide reference for the application of AI technology in production, and help improve the safety and efficiency of industrial production.

## 2. Application background of artificial intelligence in fault management

In industrial production, fault management aims at "early warning, accurate positioning, and rapid disposal", while covering three dimensions: safety, efficiency, and cost. Traditional fault management mainly relies on post-failure maintenance and regular preventive maintenance, which not only leads to long downtime and is prone to over-maintenance or insufficient maintenance, but also fault detection and diagnosis depend on personnel experience, with limited ability to identify hidden faults of complex equipment, and the results are affected by subjective factors. With the participation of artificial intelligence, multi-dimensional data is collected by sensors, industrial cameras, and IoT devices, then processed and cleaned by big data technology, and finally fault detection, diagnosis, and prediction are completed through machine learning and deep learning models, and executable decision suggestions are output, effectively making up for the shortcomings of traditional methods.

## 3. Core technologies and application cases of artificial intelligence in fault data collection and analysis

### 3.1. Image analysis for power equipment inspection based on computer vision

Computer vision simulates the human visual system through cameras, industrial cameras, and other devices to collect visual data and realize functions such as target detection, defect identification, and feature extraction. Its specific application in power inspection is mainly the target recognition and defect detection of monitoring images, which mainly combines the inspection images and videos of UAVs, inspection robots, and other equipment to identify defects in the appearance of transmission lines, towers, substations, etc. [1].

DOI: 10.54254/2755-2721/2026.BJ31086

### 3.2. Equipment predictive maintenance: a case study of the siepa-plant predictive maintenance system

Siemens has launched SiePA, a predictive maintenance software based on industrial big data analysis, also known as EPA (Equipment Predictive Analytics). Based on the in-depth analysis of historical operation data and with the help of artificial intelligence algorithms such as machine learning and natural language processing, a predictive maintenance system is established.

With the support of Siemens' artificial intelligence technology, Siemens SiePA makes full use of the factory's historical data. Through its equipment operation state prediction and early warning module and intelligent troubleshooting and diagnosis module, it can not only timely predict and warn of fault risks in operation but also help efficiently diagnose the underlying causes to guide maintenance, helping enterprises effectively control risks and achieve cost reduction and efficiency improvement.

### 3.3. Application of deep learning in the nuclear power industry

The nuclear power industry chain can be divided into upstream, midstream, and downstream parts. In the upstream nuclear fuel supply, deep learning methods can be used to construct knowledge graphs related to natural uranium, thereby improving exploration efficiency and shortening mining time. In the nuclear material processing link, deep learning methods can also be introduced to manage and analyze massive processing data to realize real-time monitoring and adjustment of processing equipment. In the midstream nuclear power equipment manufacturing, deep learning methods can be used to analyze and process massive structured, unstructured, and semi-structured data generated during the design, production, and operation of nuclear islands, conventional islands, auxiliary systems, and instrumentation and control equipment, and then provide an intelligent analysis and decision-making system. In the downstream nuclear power plant design, construction, operation, and maintenance, deep learning methods can be introduced to build an integrated, digital, and intelligent full-life-cycle nuclear power plant platform [2].

## 4. Current challenges

### 4.1. Data issues for training models

Training artificial intelligence models requires a large amount of data, and the obtained data generally has problems such as difficulty guaranteeing data quality, data interoperability and sharing issues, and unbalanced data set classification [3]. To solve these data problems, a systematic process needs to be constructed. For data quality, strict data cleaning can be used to ensure data quality and achieve "high-quality input". The purpose of data cleaning (data purification) is to detect and eliminate errors and inconsistencies in data to improve data quality [4]. For data interoperability and sharing, on the basis of formulating unified standards, federated learning can be adopted to realize "data available but not visible", breaking data silos and meeting compliance requirements. Federated learning means that training data is stored distributedly on mobile devices, and a shared model is trained by aggregating updates calculated locally [5]. For classification imbalance, a two-pronged approach is needed: using SMOTE [6] resampling at the data level, adjusting class weights or selecting models insensitive to imbalance at the algorithm level, and abandoning accuracy and adopting more effective evaluation indicators such as F1-Score and AUC-PR [7]. Finally, integrating

DOI: 10.54254/2755-2721/2026.BJ31086

these measures into a continuous iterative data governance process can lay a solid and reliable data foundation for the model.

### 4.2. Annotation cost and accuracy

In the application of artificial intelligence in fault data collection and fault analysis, it faces the core problems of high annotation cost and insufficient accuracy. The former stems from the fact that fault data is mostly high-dimensional and unstructured, requires professional knowledge for annotation, and the scarcity of fault samples will further increase the cost. The latter is due to subjective deviations of annotators, professional threshold limitations, and data noise interference, leading to inconsistent or misjudged annotation results. To solve the problems of high annotation cost and poor accuracy of artificial intelligence in fault diagnosis, the core lies in the use of machine learning. Active learning is used to select the most informative fault samples for priority annotation to avoid resource waste [8]. At the same time, semi-supervised learning is combined to fully tap a large amount of easily available unlabeled normal operation data, and physical model-based simulation data or generative AI synthetic fault samples are used to make up for data scarcity [9].

### 4.3. "Black box" dilemma

The "black box" dilemma faced by artificial intelligence in fault diagnosis means that although complex models such as deep learning can achieve high-precision prediction, their internal decision-making logic is opaque and non-traceable, making it difficult for domain experts to understand and trust the diagnosis results. In safety-critical industrial scenarios, engineers cannot confirm whether the model's judgment is based on real physical laws or pseudo-correlations in data, which not only hinders fault root cause analysis but also makes it difficult for the model to pass safety certification and integrate into the actual decision-making process, becoming a core bottleneck restricting its large-scale application.

## 5. Future development trends

The future development trends of artificial intelligence in the field of fault data collection and fault analysis can be roughly divided into four directions: developing data-efficient utilization technologies, promoting the deep integration of physical models and AI algorithms, establishing continuous learning and self-evolution mechanisms, and strengthening interdisciplinary collaboration. The aim is to systematically solve the current core problems, such as data bottlenecks, insufficient credibility, and limited adaptability, and promote industrial fault management to move towards an advanced stage of active prevention and autonomous decision-making.

### 5.1. Solving the sample insufficiency problem through meta-learning

At present, mainstream few-shot learning methods can be divided into three categories: generative learning, metric learning, and meta-learning. Among them, meta-learning is a machine learning technology that refines common knowledge across tasks and builds a general learning framework, enabling the model to quickly adapt to new tasks with only a small number of samples to achieve efficient learning and accurate decision-making. Through meta-learning, large AI models can quickly adapt to new fault samples with MAML and Relation Network algorithms [10], reduce dependence on labeled data, and improve small-sample diagnosis capabilities to solve the problem of insufficient data volume.

DOI: 10.54254/2755-2721/2026.BJ31086

### 5.2. Using prior-empowered explainable AI (XAI) to get rid of the "black box" dilemma

To get rid of the "black box" dilemma, prior-empowered explainable AI (XAI) embeds signal processing knowledge (such as wavelet transform, and time-frequency analysis) into the network structure or integrates physical laws (such as Maxwell's equations, and thermodynamic constraints) into the loss function so that the model's decision-making process is consistent with industrial mechanisms, which not only improves diagnosis accuracy but also makes fault causes traceable. For example, in gear fault diagnosis, Physics-Informed Neural Networks (PINN) significantly improve the diagnostic credibility in data-scarce scenarios through the constraints of physical loss functions [11].

### 5.3. Continuous learning and self-evolution mechanisms

In the future development of artificial intelligence in the field of fault data collection and analysis, continuous learning and self-evolution mechanisms will become the core driving force to realize truly autonomous intelligent operation and maintenance. This mechanism enables the diagnosis system to learn and evolve like humans, continuously learning from streaming data generated by equipment operation after initial deployment, dynamically adapting to the natural degradation of equipment, working condition fluctuations, and new fault modes without manual retraining. Specifically: through continuous learning technologies (such as elastic weight consolidation, dynamic network expansion), the system avoids catastrophic forgetting of already mastered fault features while absorbing new knowledge; through the meta-learning framework, the model gains the ability of "learning to learn" and can quickly adapt when facing a small number of new fault samples; further, combined with automated machine learning and neural architecture search, the system will evolve into a "self-evolving" entity that can self-optimize its internal structure and independently adjust diagnosis strategies, ultimately realizing the leap from a static diagnostic tool to a dynamic intelligent operation and maintenance partner, improving human-machine collaboration capabilities, and providing full-life-cycle adaptive health management for industrial systems [12].

### 5.4. Construction of a standardized system for interdisciplinary integration

In the future, with the gradual improvement of the precision requirements for fault data collection and fault analysis, it is necessary to promote the cross-integration of artificial intelligence with multiple disciplines, such as electronic and information engineering, computer science, industrial engineering, and control engineering. Interdisciplinary collaboration is supported by AI as the core, and the construction of a standardized system is the top priority of cross-integration. At the technical standard level, full-process specifications covering data collection, model evaluation, and diagnosis result certification will be established, unifying the interface formats of multi-source data such as vibration signals and visual data, and formulating quantitative indicators for model interpretability and robustness. At the interdisciplinary collaboration level, the deep integration of industrial engineering, data science, and physics will become the norm: mechanical engineers provide equipment mechanism priors, data scientists optimize algorithm architectures, and jointly develop special diagnostic systems suitable for specific scenarios, such as a full-life-cycle monitoring platform in the nuclear power field that integrates nuclear physics knowledge and deep learning. In addition, the "enterprise-research institution" collaborative innovation model will accelerate

DOI: 10.54254/2755-2721/2026.BJ31086

technology transformation and promote the large-scale application of standardized products such as predictive maintenance software and intelligent inspection equipment [13].

## 6. Conclusion

This study systematically reviews the application value, core achievements, and existing challenges of artificial intelligence technology in the field of industrial fault data collection and fault analysis, providing comprehensive theoretical reference and practical guidance for the intelligent development of industrial fault management. The research indicates that at the critical stage of the intelligent transformation under Industry 4.0, AI technology has demonstrated irreplaceable advantages in fault management across multiple industries through core methods such as computer vision, deep learning, and machine learning: in power equipment inspection, accurate defect identification is achieved, greatly enhancing inspection efficiency and safety; predictive maintenance systems like Siemens SiePA enable fault risk warning and root cause diagnosis by mining historical operational data, helping enterprises reduce costs and improve efficiency; deep learning technology runs through the entire nuclear power industry chain, constructing a full-life-cycle intelligent management system. These applications have effectively compensated for the shortcomings of traditional fault management models, significantly improving operational efficiency and safety.

However, the current application of AI in fault data still faces many challenges: at the data level, issues such as uneven data quality, difficulties in cross-subject data sharing, and imbalanced dataset classifications exist; the annotation process suffers from high costs and difficulties in ensuring accuracy; the "black-box" nature of models leads to insufficient decision credibility, and their dynamic adaptability to equipment degradation and fluctuating working conditions remains limited. These intertwined problems form the core bottlenecks hindering the in-depth implementation and large-scale promotion of AI technology.

Future research should focus on the following directions: developing data-efficient utilization technologies like few-shot learning, combined with generative AI and digital twins to address data scarcity; promoting the deep integration of physical models and AI algorithms to enhance model interpretability and credibility; establishing continuous learning and self-evolution mechanisms for dynamic model optimization; and strengthening interdisciplinary collaboration to formulate unified standards and a full-process specification system. Through these improvements, AI-enabled fault management will advance towards a more autonomous, reliable, and efficient stage, injecting sustained momentum into industrial intelligent transformation and fostering the development of a safe, efficient, and green modern industrial production ecosystem.

## References

[1] Pu T J, Qiao J, Han X, Zhang G B, Wang X Y. (2020). Research and application of artificial intelligence technology in operation, maintenance and overhaul of power equipment. High Voltage Engineering, 46(02), 369-383.
[2] C. Tang et al. (2022) Deep learning in nuclear industry: A survey. Big Data Mining and Analytics, June, vol. 5, no. 2, pp. 140-160. doi: 10.26599/BDMA.2021.9020027.
[3] Tang W. H., Niu Z. W., Zhao B. N., Ji T. Y., Li M. S., Wu Q. H. (2020). Research and application of data-driven artificial intelligence technology in power equipment condition analysis. High Voltage Engineering, 46(09), 2985-2999.
[4] Li Cai & Yangyong Zhu. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14, 2-2. https: //doi.org/10.5334/dsj-2015-002.
[5] Mcmahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, Blaise Agüera y. (2016). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial

DOI: 10.54254/2755-2721/2026.BJ31086

Intelligence and Statistics (AISTATS) Fort Lauderdale, Florida, USA. JMLR: W&CPvolume 54.
[6] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall & W. Philip Kegelmeyer. (2002). SMOTE: Synthetic Minority Over-sampling Technique. The Journal of Artificial Intelligence Research, 16(0), 321-357.
[7] Powers D.M.W. (2011). Evaluation: From Precision, Recall and F-Measure to Roc, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies, 2(1), 37-63.
[8] Settles, B. (2009). Active Learning Literature Survey. University of Wisconsin–Madison. https: //minds.wisconsin.edu/handle/1793/60660
[9] Zhu, X. J. (2005). Semi-Supervised Learning Literature Survey. University of Wisconsin-Madison. http: //digital.library.wisc.edu/1793/60444
[10] Likui Qiao, Yuxian Zhang, Qisen Wang, Donglin Li & Shidong Peng. (2025). Fault diagnosis for wind turbine generators based on Model-Agnostic Meta-Learning: A few-shot learning method. Expert Systems with Applications, 267, 126171-126171.
[11] Yan R. Q., Shang Z. G., Wang Z. Y., Xu W. G., Zhao Z. B., Wang S. B., Chen X. F. (2024). Challenges and opportunities of explainable artificial intelligence in industrial intelligent diagnosis: Prior empowerment. Journal of Mechanical Engineering, 60(12), 1-20.
[12] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, Stefan Wermter. (2019). Continual lifelong learning with neural networks: A review. Neural Networks, 113, 54-71. https: //doi.org/10.1016/j.neunet.2019.01.012.
[13] Liu X. Y., Sun L. N., Qiu J. W., Shan X. H. (2022). Cross-border technology integration in the field of artificial intelligence based on multi-layer networks. Complex Systems and Complexity Science, 19(01), 45-51.